

Blind Source Separation with Compressively Sensed Linear Mixtures¹

Martin Kleinstaubert and Hao Shen

October 13, 2011

Abstract

This work studies the problem of simultaneously separating and reconstructing signals from compressively sensed linear mixtures. We assume that all source signals share a common sparse representation basis. The approach combines classical Compressive Sensing (CS) theory with a linear mixing model. It allows the mixtures to be sampled independently of each other. If samples are acquired in the time domain, this means that the sensors need not be synchronized. Since Blind Source Separation (BSS) from a linear mixture is only possible up to permutation and scaling, factoring out these ambiguities leads to a minimization problem on the so-called oblique manifold. We develop a geometric conjugate subgradient method that scales to large systems for solving the problem. Numerical results demonstrate the promising performance of the proposed algorithm compared to several state of the art methods.

Index Terms

Compressed sensing, blind source separation, oblique manifold, conjugate subgradient method.

1 Introduction

Recovering signals from only the mixed observations without knowing the priori information of both the source signals and the mixing process is often referred to as *Blind Source Separation* (BSS), cf. [2]. Different BSS methods are used in various challenging data analysis applications, such as functional Magnetic Resonance Imaging (fMRI) analysis and microarray analysis. In order to achieve reasonable performance, prominent methods, e.g. Independent Component Analysis (ICA), usually require a large number of observations [3]. Unfortunately, the availability of a large amount of data samples can not be guaranteed in many real applications, due to either cost or time issues.

The theory of compressed sensing (CS), cf. [4], shows that, when a signal is *sparse* (or *compressible*) with respect to some basis, only a small number of samples suffice for exact (or *approximate*) recovery. It is interesting to know that the concept of sparsity has also been used as a separation criterion in the context

¹M. Kleinstaubert and H. Shen are with the Department of Electrical Engineering and Information Technology, Technische Universität München, München, Germany. e-mail: (see <http://www.gol.ei.tum.de>).

Authors are listed in alphabetical order due to equal contribution.

Parts of this work have been presented at the Workshop *Signal Processing with Adaptive Sparse Structured Representations (SPARS'11)*, Edinburgh, June 2011 [1].

This work has partially been supported by the Cluster of Excellence *CoTeSys - Cognition for Technical Systems*, funded by the Deutsche Forschungsgemeinschaft (DFG).

of BSS. In [5], for example, a family of efficient algorithms in the probabilistic framework are proposed. In [6], the authors investigate sparse representation of signals in an overcomplete basis in conjunction with BSS and motivate the usage of ℓ_1 -norm for a sparsity measure.

The approach presented in this work can be regarded as a generalization of Morphological Component Analysis (MCA), cf. [7] and [8]. Roughly speaking, while MCA takes advantage of the sparse representations of the signals only for separation tasks, we additionally employ the sparsity assumption for reconstruction. The resulting cost function is thus very much related to the ones that typically arise in MCA. Nevertheless, the current scenario with compressively sensed samples has not been studied and thus differs from the existing MCA models. Although it has the potential of tackling the underdetermined BSS case, our numerical validation only considers the determined case, where the number of sources equals the number of mixtures.

By considering the classical result that the source signals can only be extracted with scaling ambiguity, we regularize our problem by restricting each column of the mixing matrix to have unit norm. In other words, the resulting optimization problem is defined on the so-called oblique manifold. Furthermore, as many potential applications of the present work lie in the area of image or audio signal processing, any promising algorithms have to be capable of scaling to large systems. It is well known that conjugate gradient-type methods provide a nice trade-off between large scale problems and good convergence behavior. In this work, we propose an approach that lifts the conjugate subgradient method proposed by Wolfe in [9] to the manifold case.

The paper is outlined as follows. Section 2 introduces some basic concepts and provides a description of the compressively sensed BSS problem. In Section 3, we develop a geometric subgradient algorithm. Finally, numerical results and conclusions are given in Section 4 and Section 5 respectively.

2 Problem Description

2.1 Notation and Setting

For the sake of convenience of presentation, we follow the notation in the compressive sensing literature and represent the signals as column vectors. The instantaneous linear BSS model is given by

$$Y = SA, \tag{1}$$

where $S = [s_1, \dots, s_m] \in \mathbb{R}^{n \times m}$ denotes the data matrix of m sources with n samples ($m \ll n$), $A = [a_1, \dots, a_k] \in \mathbb{R}^{m \times k}$ is the mixing matrix of full rank, and $Y = [y_1, \dots, y_k] \in \mathbb{R}^{n \times k}$ represents the k linear mixtures of S . The task of standard BSS is to estimate the sources S , given only the mixtures Y .

We assume that for all sources $s_i \in \mathbb{R}^n$, there exists an (overcomplete) basis $\mathcal{D} \in \mathbb{R}^{n \times d}$, $n \leq d$, of \mathbb{R}^n , referred to as *representation basis*, such that $s_i = \mathcal{D}x_i$, with sparse $x_i \in \mathbb{R}^d$. More compactly, this reads as

$$S = \mathcal{D}X, \tag{2}$$

where $X = [x_1, \dots, x_m] \in \mathbb{R}^{d \times m}$.

2.2 Compressively sensed BSS model

Now let us take one step further to compressively sample each mixture $y_i \in \mathbb{R}^n$ individually. To this end, we denote the *sampling matrix* for the mixture y_i by $\Phi_i \in \mathbb{R}^{p_i \times n}$ for $i = 1, \dots, k$. Then, a compressively sensed observation $\hat{y}_i \in \mathbb{R}^{p_i}$ of the i -th mixture is given by

$$\hat{y}_i = \Phi_i y_i = \Phi_i \mathcal{D} X a_i. \quad (3)$$

We refer to (3) as the *Compressively Sensed BSS* (CS-BSS) model. To summarize, the task considered in this work is formulated as: Given the compressively sensed observations $\hat{y}_i \in \mathbb{R}^{p_i}$, for $i = 1, \dots, k$, together with their corresponding sampling matrices $\Phi_i \in \mathbb{R}^{p_i \times n}$, estimate the mixing matrix $A \in \mathbb{R}^{m \times k}$ and the sparse representations $X \in \mathbb{R}^{n \times m}$.

There are various measures of sparsity available in the literature, cf. [10]. In this work, we confine ourselves to the ℓ_1 -norm as a measure of sparsity, because (i) it is suitable for numerical algorithms, in particular for large scale problems; (ii) it is an appropriate prior for many real signals, cf. [5]. This leads to the following optimization problem

$$\underset{X, A}{\operatorname{argmin}} \|X\|_1, \quad \text{s.t.} \quad \hat{y}_i = \Phi_i \mathcal{D} X a_i, \quad i = 1, \dots, k. \quad (4)$$

In real applications, it is unavoidable that the observations \hat{y}_i are contaminated by noise. Hence let ϵ_i denote the error radius for the observation i . The optimization problem (4) turns into

$$\underset{X, A}{\operatorname{argmin}} \|X\|_1, \quad \text{s.t.} \quad \|\hat{y}_i - \Phi_i \mathcal{D} X a_i\|_2 \leq \epsilon_i, \quad i = 1, \dots, k. \quad (5)$$

Following a standard approach, we formulate (5) in an unconstrained Lagrangian form, namely

$$\underset{X, A}{\operatorname{argmin}} \|X\|_1 + \sum_{i=1}^k \lambda_i \|\hat{y}_i - \Phi_i \mathcal{D} X a_i\|_2^2, \quad (6)$$

where the scalars $\lambda_i \in \mathbb{R}^+$ weigh the reconstruction error of each mixture individually according to $\lambda_i \sim 1/\epsilon_i^2$, and balance these errors against the sparsity term $\|X\|_1$. It is well known that, in compressive sensing, inappropriate regularization parameters might lead to not only slow convergence but also local optima. To cope with this issue, we follow the adaptive updating strategy proposed in [11, 12]. A rigorous analysis of the updating strategy in our CS-BSS setting is beyond the scope of this work.

2.3 Regularized CS-BSS problem

Obviously, problem (6) is ill-posed. Indeed, an optimization procedure would let the norm of A explode to drive $\|X\|_1$ to zero. In order to regularize the problem, we therefore restrict to the case where $\|a_i\|_2 = 1$ for all $i = 1, \dots, k$. Thus, together with the full rank condition, we restrict the mixing matrix A onto the oblique manifold [13]

$$\mathcal{OB}(m, k) := \{A \in \mathbb{R}^{m \times k} \mid \operatorname{rk}(A) = k, \operatorname{ddiag}(A^\top A) = I_k\}, \quad (7)$$

where I_k is the $(k \times k)$ identity matrix, and $\text{ddiag}(Z)$ forms a diagonal matrix, whose diagonal entries are those of Z . Note, that this is a common approach in many BSS scenarios, since it is well known [14] that the mixing matrix A is identifiable only up to a column-wise scaling and permutation. The regularized problem that we will consider is hence given by

$$\underset{\substack{X \in \mathbb{R}^{d \times m} \\ A \in \mathcal{OB}(m, k)}}{\text{argmin}} \quad \|X\|_1 + \sum_{i=1}^k \lambda_i \|\hat{y}_i - \Phi_i \mathcal{D}X a_i\|_2^2. \quad (8)$$

The optimization problem is thus defined on the product manifold $\mathbb{R}^{d \times m} \times \mathcal{OB}(m, k)$.

3 A Conjugate Subgradient type method for the CS-BSS problem

3.1 Conjugate Subgradient on Riemannian manifolds

In order to tackle the non-smooth problem (8), we propose a conjugate subgradient type approach on manifolds by generalizing the method proposed in [9]. A generalization of subgradient methods to the Riemannian manifold case has been studied in [15]. We refer to [16] for an introduction to nonsmooth analysis on manifolds, where the required concepts are introduced. Generally, for minimizing a nonsmooth function $f: M \rightarrow \mathbb{R}$ where M is a Riemannian manifold and a subgradient of f exists for all $x \in M$, we propose the following scheme, further referred to as *Riemannian Conjugate Subgradient (CSG) method*. Let $T_x M$ be the tangent space at x , let $\partial f(x) \subset T_x M$ denote the Riemannian sub-differential of f at x , i.e. the set of all Riemannian subgradients at x , and let $\|\cdot\|_R$ denote the norm on $T_x M$ which is induced by the Riemannian metric $\langle \cdot, \cdot \rangle_R$. Moreover, let

$$G(x) := \underset{\xi \in \partial f(x)}{\text{argmin}} \|\xi\|_R \quad (9)$$

denote the subgradient with smallest Riemannian norm. Let $x_0 \in M$ be an initial point for the algorithm. Denote $G_i := G(x_i)$, and the descent direction at the i -th iteration by $H_i \in T_{x_i} M$. By initializing $H_0 := -G_0$, the algorithm now iteratively updates

$$x_{i+1} = \exp_{x_i}(\alpha_i H_i), \quad (10)$$

where \exp denotes the Riemannian exponential and the scalar α_i is the line search parameter at the i -th iteration. Various line-search techniques for computing α_i exist from which we choose *backtracking line-search* [17], cf. also [18] for the Riemannian case, as it is conceptually simple and computationally cheap.

There are also several formulas

available that update the descent direction, e.g. Hestenes-Stiefel, Polak-Ribière, and Fletcher-Reeves, which can all be generalized straightforwardly to the manifold setting. Here, we restrict to a manifold adaption of Hestenes-Stiefel. Let $\xi \in T_{x_i} M$ and denote by $\tau_i(\xi) \in T_{x_{i+1}} M$ the parallel transport of

ξ along the geodesic $\gamma(t) = \exp_{x_i}(tH_i)$ into the tangent space $T_{x_{i+1}}M$. The direction update is then given by

$$H_{i+1} := -G_{i+1} + \frac{\langle G_{i+1}, G_{i+1} - \tau_i(G_i) \rangle_R}{\langle \tau_i(H_i), G_{i+1} - \tau_i(G_i) \rangle_R} \tau_i(H_i). \quad (11)$$

Equations (10) and (11) are iterated until a stopping criterion is met. Usually, conjugate gradient methods use a reset after $N = \dim M$ iterations, i.e. $H_i := -G_i$ if $i \bmod N = 0$. We propose to reset the direction when no significant update on the search directions is observed.

3.2 The Riemannian subgradient for CS-BSS

The manifold $M = \mathbb{R}^{d \times m} \times \mathcal{OB}(m, k)$ is endowed with the Riemannian metric inherited from the surrounding Euclidean space, that is

$$\langle (X_1, A_1), (X_2, A_2) \rangle_R := \text{tr}(X_1 X_2^\top) + \text{tr}(A_1 A_2^\top). \quad (12)$$

Let $\lambda_i > 0$ for $i = 1, \dots, k$ and let

$$f: \mathbb{R}^{d \times m} \times \mathcal{OB}(m, k) \rightarrow \mathbb{R},$$

$$f(X, A) = \|X\|_1 + \sum_{i=1}^k \lambda_i \|\hat{y}_i - \Phi_i \mathcal{D}X a_i\|_2^2. \quad (13)$$

The Riemannian subdifferential of f at (X, A) with respect to the Riemannian metric (12) is given by

$$\partial f(X, A) = \{(\partial_1 f(X, A), \partial_2 f(X, A))\}, \quad (14)$$

where

$$\partial_1 f(X, A) = \partial \|X\|_1 + \sum_{i=1}^k \lambda_i (\Phi_i \mathcal{D})^\top (\Phi_i \mathcal{D}X a_i - \hat{y}_i) a_i^\top \quad (15)$$

and

$$\partial_2 f(X, A) = [\lambda_i \Pi(a_i) (\Phi_i \mathcal{D}X)^\top (\Phi_i \mathcal{D}X a_i - \hat{y}_i)]_{i=1, \dots, k}. \quad (16)$$

Here, $\partial \|X\|_1 = \{H \in \mathbb{R}^{d \times m} \mid H_{ij} = \text{sign}(X_{ij}) \text{ if } X_{ij} \neq 0, H_{ij} \in [-1, 1] \text{ else}\}$, and $\Pi(a_i) := I_m - a_i a_i^\top$ is an orthogonal projection operator.

Note, that $\partial_2 f(X, A)$ is in fact a Riemannian gradient on the oblique manifold. Hence we may write $\partial_2 f(X, A) = \nabla_2 f(X, A)$. For the purpose of finding the subgradient with smallest Riemannian norm, each entry of $\partial_1 f(X, A)$ can be minimized independently. Let us denote the second summand at the right-hand side of (15) by

$$B := \sum_{i=1}^k \lambda_i \mathcal{D}^\top \Phi_i^\top (\Phi_i \mathcal{D}X a_i - \hat{y}_i) a_i^\top \quad (17)$$

and define $C \in \mathbb{R}^{d \times m}$ as the matrix with (i, j) -entries

$$C_{ij} := \begin{cases} \text{sign}(X_{ij}) & \text{if } X_{ij} \neq 0 \\ -\text{sign}(B_{ij}) \min\{|B_{ij}|, 1\} & \text{otherwise.} \end{cases} \quad (18)$$

Then the subgradient $G(X, A) \subset \partial f(X, A)$ having smallest Riemannian norm is given explicitly by

$$G(A, X) = (C + B, \nabla_2 f(X, A)). \quad (19)$$

3.3 A Conjugate Subgradient Algorithm

The development of a Riemannian Conjugate Subgradient method requires the concept of parallel transport along a geodesic. Since we consider the oblique manifold as a Riemannian submanifold of a product of unit spheres, the formulas for the parallel transport as well as the exponential mapping read accordingly.

Let

$$\tau_{x,\xi}(\psi) := \psi - \frac{\xi^\top \psi}{\|\xi\|^2} (x\|\xi\| \sin t\|\xi\| + \xi(1 - \cos t\|\xi\|)). \quad (20)$$

be the parallel transport of a tangent vector $\psi \in T_x S^{m-1}$ along the great circle $\gamma_{x,\xi}(t)$ in the direction $\xi \in T_x S^{m-1}$ on the unit sphere S^{m-1} , where

$$\begin{aligned} \gamma_{x,\xi}: \mathbb{R} &\rightarrow S^{m-1}, \\ \gamma_{x,\xi}(t) &:= \begin{cases} x, & \|\xi\| = 0; \\ x \cos t\|\xi\| + \xi \frac{\sin t\|\xi\|}{\|\xi\|}, & \text{otherwise.} \end{cases} \end{aligned} \quad (21)$$

Then the parallel transport of $(H, \Psi) \in T_{(X,A)} \mathbb{R}^{d \times m} \times \mathcal{OB}(m, k)$ with respect to the Levi-Civita connection along the geodesic $\gamma_{(X,A)}$ in the direction $(Z, \Xi) \in T_{(X,A)} \mathbb{R}^{d \times m} \times \mathcal{OB}(m, k)$, i.e.

$$\begin{aligned} \gamma_{(X,A)}: \mathbb{R} &\rightarrow \mathbb{R}^{d \times m} \times \mathcal{OB}(m, k), \\ \gamma_{(X,A)}(t) &:= (X + tZ, [\gamma_{a_i, \xi_i}(t)]_{i=1, \dots, k}) \end{aligned} \quad (22)$$

is given by

$$\tau_{(X,A),(Z,\Xi)}(H, \Psi) := \left(H, [\tau_{a_i, \xi_i}(\psi_i)]_{i=1, \dots, k} \right). \quad (23)$$

A conjugate subgradient algorithm for the problem as defined in (8) is summarized as follows.

Algorithm 1 *A conjugate subgradient CS-BSS algorithm*

Step 1: Initialize $X^{(0)} \in \mathbb{R}^{d \times m}$ and $A^{(0)} \in \mathcal{OB}(m, k)$.

Set $i = 0$.

Step 2: Set $i = i + 1$, let $X^{(i)} = X^{(i-1)}$ and $A^{(i)} = A^{(i-1)}$, and compute

$$G^{(1)} = H^{(1)} = -G(A^{(i)}, X^{(i)}).$$

Step 3: For $j = 1, \dots, d \cdot m + m(m-1) - 1$:

(i) Update $(X^{(i)}, A^{(i)}) \leftarrow \gamma_{(X^{(i)}, A^{(i)}), H^{(i)}}(\lambda^*)$,

where

$$\lambda^* = \operatorname{argmin}_{\lambda \in \mathbb{R}} \gamma_{(X^{(i)}, A^{(i)}), H^{(i)}}(\lambda);$$

(ii) Compute $G^{(j+1)} = -G(A^{(i)}, X^{(i)})$;

(iii) Update $H^{(j+1)} \leftarrow G^{(j+1)} + \mu \tau_{(X^{(i)}, A^{(i)}), H^{(j)}}(H^{(j)})$, where μ is chosen according to (11);

(iv) If $\|H^{(j+1)} - H^{(j)}\|$ is small enough, go to Step 2.

Step 4: If $\|A^{(i+1)} - A^{(i)}\| + \|X^{(i+1)} - X^{(i)}\|$ is small enough, stop. Otherwise, go to Step 2.

4 Numerical Results

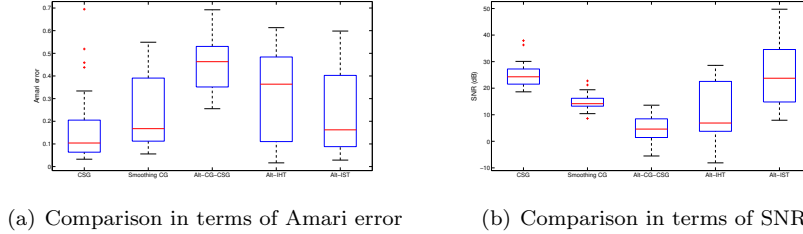


Figure 1: Experimental results.

In our experiment, we investigate the performance of the proposed CSG algorithm compared with several state of the art algorithms. By noticing the similarity between our present work with the generalized MCA method in [8], we adapt the alternating soft-thresholding based method and the hard thresholding [19] (where ℓ_0 is used as a sparsity measure) to our CS-BSS setting. We refer to them as *Alt-IST* and *Alt-IHT*, respectively. Similarly, we adapt the CSG algorithm in the alternating manner (referred to as *Alt-CSG-CSG*). Finally, one alternative common approach to deal with the nondifferentiability of the cost function is to employ a smooth approximation. We then refer to the standard CG algorithm with certain smooth approximation as *smoothing CG*.

In the experiment, we generate three sources signals with $n = 1000$ samples. The number of nonzero entries is equal to 100, which are generated according to a Laplacian distribution. We randomly pick 300 samples as the compressively sensed mixtures. The performance is measured in terms of both separation quality, measured by the Amari Error [20], and reconstruction quality in terms of Signal to Noise Ratio (SNR). Figure I illustrates the box plot of the results from 50 experiments of all methods. Among all five methods, the *Alt-CSG-CSG* performs the worst on average in terms of both separation and reconstruction qualities, while the proposed CSG algorithm provides consistent and satisfactory results. Finally, the small variances of all CG/CSG type algorithms in terms of reconstruction also suggest their reliable performance compared with the thresholding based approaches.

5 Conclusion

In this work, the authors propose a method for separating linearly mixed sparse signals from compressively sensed mixtures. The proposed approach allows each mixture to be sensed individually. If sampling takes place in the time domain, this means that the sensors do not have to be synchronized. The arising optimization problem is tackled with a conjugate subgradient type method, which is based on a new concept for optimizing non-differentiable functions under smooth constraints.

References

- [1] M. Kleinstaubler and H. Shen, “Blind source separation of compressively sensed signals,” in *Proceedings of the 4th Workshop on Signal Processing with Adaptive Sparse Structured Representations*, 2011, p. 80.
- [2] S. Haykin, *Unsupervised Adaptive Filtering*. Wiley-Interscience, 2000, vol. 1: Blind Source Separation.
- [3] S. Bermejo, “Finite sample effects in higher order statistics contrast functions for sequential blind source separation,” *IEEE Signal Processing Letters*, vol. 12, no. 6, pp. 481–484, 2005.
- [4] D. L. Donoho, “Compressed sensing,” *IEEE Transactions on Information Theory*, vol. 52, pp. 1289–1306, 2006.
- [5] M. Zibulevsky and B. A. Pearlmutter, “Blind source separation by sparse decomposition in a signal dictionary,” *Neural Computation*, vol. 13, no. 4, pp. 863–882, 2001.
- [6] Y. Li, A. Cichocki, and S.-i. Amari, “Analysis of sparse representation and blind source separation,” *Neural Computation*, vol. 16, pp. 1193–1234, 2004.
- [7] M. Elad, J. Starck, P. Querre, and D. L. Donoho, “Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA),” *Applied and Computational Harmonic Analysis*, vol. 19, no. 3, pp. 340–358, 2005.
- [8] J. Bobin, J.-L. Starck, J. Fadili, and Y. Moudden, “Sparsity and morphological diversity in blind source separation,” *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2662–2674, 2007.
- [9] P. Wolfe, “A method of conjugate subgradients for minimizing nondifferentiable functions,” in *Nondifferentiable Optimization*, ser. Mathematical Programming Studies, 1975, vol. 3, pp. 145–173.
- [10] N. Hurley and S. Rickard, “Comparing measures of sparsity,” *IEEE Transactions on Information Theory*, vol. 55, no. 10, pp. 4723–4741, 2009.
- [11] J. Bobin, J. Starck, J. M. Fadili, Y. Moudden, and D. L. Donoho, “Morphological component analysis: An adaptive thresholding strategy,” *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2675–2681, 2007.
- [12] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo, “Sparse reconstruction by separable approximation,” *IEEE Transactions on Signal Processing*, vol. 57, pp. 2479–2493, 2009.
- [13] H. Shen and M. Kleinstaubler, “Complex blind source separation via simultaneous strong uncorrelating transform,” in *Lecture Notes in Computer Science, Proceedings of the 9th International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA 2010)*, vol. 6365. Berlin/Heidelberg: Springer-Verlag, 2010, pp. 287–294.

- [14] P. Comon, “Independent component analysis, a new concept?” *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.
- [15] O. P. Ferreira and P. R. Oliveira, “Subgradient algorithm on Riemannian manifolds,” *Journal of Optimization Theory and Applications*, vol. 97, pp. 93–104, 1998.
- [16] Y. S. Ledyev and Q. J. Zhu, “Nonsmooth analysis on smooth manifolds,” *Transactions of the American Mathematical Society*, vol. 359, no. 8, pp. 3687–3732, 2007.
- [17] J. Nocedal and S. J. Wright, *Numerical Optimization, 2nd Ed.* New York: Springer, 2006.
- [18] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds.* Princeton, NJ: Princeton University Press, 2008.
- [19] T. Blumensath and M. E. Davies, “Iterative hard thresholding for compressed sensing,” *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265–274, 2009.
- [20] S. Amari, A. Cichocki, and H. H. Yang, “A new learning algorithm for blind signal separation,” *Advances in Neural Information Processing Systems*, vol. 8, pp. 757–763, 1996.